

UWAGA, ODPOWIEDZI W KOLEJNOŚCI LOSOWEJ, KLUCZ NA KOŃCU

Imię i nazwisko:

Nr indeksu:

1. W pewnym korpusie 10. słowo wg frekwencji miało częstość 100000, jak będzie oczekiwana częstość 40. słowa?
  - 2500
  - 100040
  - 50000
  - 25000
2. Po zastosowaniu truecasingu "LECH POZNAŃ WITA SERDECZNIE GOŚCI Z WARSZAWY!" przyjmie postać:
  - lech Poznań wita serdecznie gości z Warszawy!
  - lech poznań wita serdecznie gości z warszawy!
  - Lech poznań wita serdecznie gości z warszawy!
  - Lech Poznań wita serdecznie gości z Warszawy!
3. Przykładowa tokenizacja zdania "Wysłuchaliśmy wykładu prof. Adama Nowaka." to:
  - Wysłuchali|śmy| |wykład|u| |prof|.| |Adam|a| |Nowak|a|.
  - Wysłuchali|śmy| |wykład|u| |prof|.| |Adam|a| |Nowak|a|.
  - Wysłuchaliśmy| |wykładu| |prof|.| |Adama Nowaka|.
  - Wysłuchaliśmy| |wykładu| |prof|.| |Adama| |Nowaka|.
4. Tłumaczymy z polskiego na angielski. Jako tłumaczenie zdania „Ala ma kota” rozpatrujemy zdanie „Alice has a cat”. Naturalna funkcja urownowleglenia (alignment function) to:
  - $1 \rightarrow 1, 2 \rightarrow 2, 3 \rightarrow 3$
  - $1 \rightarrow 1, 2 \rightarrow 2, 3 \rightarrow 4$
  - $1 \rightarrow 1, 2 \rightarrow 2, 3 \rightarrow \text{NULL}, 4 \rightarrow 3$
  - $1 \rightarrow 1, 2 \rightarrow 2, 3 \rightarrow 2, 4 \rightarrow \text{NULL}$
5. Tłumaczymy z polskiego na angielski. Zdanie polskie ma 3 wyrazy, zaś zdanie angielskie — 2 wyrazy. Ile jest możliwych funkcji urownowleglenia (niekoniecznie zgodnych z sensem)?
  - 27
  - 9
  - 16
  - 8
6. Załóżmy, że  $P(\text{rach}|\text{foo})=0.5$ ,  $P(\text{ciach}|\text{foo})=0.5$ ,  $P(\text{ciach}|\text{NULL})=0.1$ ,  $P(\text{rach}|\text{bar})=0.7$ ,  $P(\text{ciach}|\text{bar})=0.3$ ,  $P(\text{rach}|\text{NULL})=0.8$ . Jakie jest najbardziej prawdopodobne dopasowanie w Modelu 1? Rozpatrujemy zdanie źródłowe "foo bar" i zdanie docelowe "rach ciach".
  - $1 \rightarrow 1, 2 \rightarrow 2$
  - $1 \rightarrow \text{NULL}, 2 \rightarrow 2$

- $1 \rightarrow \text{NULL}, 2 \rightarrow 1$
  - $1 \rightarrow 2, 2 \rightarrow 1$
7. Rozpatrzmy korpus złożony z dwóch par zdań: „foo bar”=„rach ciach” i „bar baz”=„rach trach”. Jeśli rozpatrywać w izolacji dopasowania: dla pierwszego zdania —  $1 \rightarrow 2, 2 \rightarrow 2$ , dla drugiego zdania —  $1 \rightarrow 1, 2 \rightarrow 2$ , jak byśmy wyliczyli prawdopodobieństwo  $P(\text{rach}|\text{bar})$ ?
- 0.75
  - 0.25
  - 1.0
  - 0.5
8. Rozpatrujemy korpus złożony z dwóch par zdań: „foo bar”=„rach ciach” i „bar baz”=„rach trach bach” (zakładamy, że w ogóle nie ma innych wyrazów). Jaka będzie wartość  $P(\text{rach}|\text{bar})$  po inicjalizacji równym prawdopodobieństwem (krok początkowy algorytmu EM)? Uwzględniamy token NULL.
- $1/3$
  - $1/4$
  - $1/2$
  - $1/5$
9. Przy zastosowaniu algorytmu EM dla Modelu 1 dla korpusu złożonego z dwóch par zdań: „foo bar”=„rach ciach” i „bar baz”=„rach trach”, jaka będzie wartość  $P(\text{rach}|\text{bar})$  po jednej iteracji algorytmu EM (inicjalizujemy równomiernym prawdopodobieństwem i uwzględniamy token NULL)?
- 0.33333
  - 1
  - 0.25
  - 0.5
10. Przy zastosowaniu algorytmu EM dla Modelu 1 dla korpusu złożonego z dwóch par zdań: „foo bar”=„rach ciach” i „bar baz”=„rach trach”, jaka będzie wartość  $P(\text{rach}|\text{bar})$  po dwóch iteracjach algorytmu EM (inicjalizujemy równomiernym prawdopodobieństwem i uwzględniamy token NULL)?
- $4/7$
  - 1
  - $1/2$
  - $3/14$
11. Przy zastosowaniu algorytmu EM dla Modelu 1 dla korpusu złożonego z dwóch par zdań: „foo bar”=„rach ciach” i „foo bar baz”=„rach trach ciach”, po wielu iteracjach algorytmu EM (inicjalizujemy równomiernym prawdopodobieństwem i uwzględniamy token NULL):
- $P(\text{rach}|\text{bar}) = 1, P(\text{trach}|\text{baz}) = 1$
  - $P(\text{rach}|\text{bar}) = 1/2, P(\text{trach}|\text{baz}) = 1/3$
  - $P(\text{rach}|\text{bar}) = 1/2, P(\text{trach}|\text{baz}) = 1$
  - $P(\text{rach}|\text{bar}) = 1/3, P(\text{trach}|\text{baz}) = 1/3$

12. Dla Modelu 1 dla korpusu złożonego z 2 par zdań: „foo bar baz”=„rach ciach trach” i „bar baz foo”=„trach ciach rach”, jaka będzie wartość  $P(\text{rach}|\text{bar})$  po wielu iteracjach algorytmu EM (inicjalizujemy równomiernym prawdopodobieństwem i uwzględniamy NULL).
- 7/8
  - 0
  - 1/2
  - 1/3
13. Rozpatrujemy korpus złożony z 2 par zdań: „foo bar baq”=„rach ciach aha” i „foo bar baz”=„rach trach ciach” (zakładamy, że w ogóle nie ma innych wyrazów). Jaka będzie wartość  $a(2|1,3,3)$  po inicjalizacji równym prawdopodobieństwem (krok początkowy algorytmu EM dla Modelu 2)? Uwzględniamy token NULL.
- 1/3
  - 1/2
  - 1/4
  - 2/3
14. Rozpatrujemy korpus złożony z 2 par zdań: „foo bar baq”=„rach ciach aha” i „foo bar baz”=„rach trach ciach” (zakładamy, że w ogóle nie ma innych wyrazów). Jaka będzie wartość  $a(2|1,3,3)$  po pierwszym kroku algorytmu EM dla Modelu 2? Uwzględniamy token NULL.
- 3/5
  - 0
  - 1/4
  - 1/3
15. Prawdopodobieństwa leksykalne w Modelu 2 inicjalizowane są:
- w sposób losowy.
  - wartościami  $1/n$ , gdzie  $n$  jest długością zdania źródłowego.
  - wartościami  $1/n$ , gdzie  $n$  jest długością zdania docelowego.
  - prawdopodobieństwami leksykalnymi wyliczonymi wg Modelu 1.
16. Symetryzacja dopasowań. Dla zdań „Mieszkańcy całej ziemi mieli jedną mowę” i „And the whole earth was of one language” mamy funkcję dopasowania: (w kierunku polsko-angielskim)  $1 \rightarrow \text{NULL}$ ,  $2 \rightarrow 2$ ,  $3 \rightarrow 2$ ,  $4 \rightarrow 3$ ,  $5 \rightarrow \text{NULL}$ ,  $6 \rightarrow \text{NULL}$ ,  $7 \rightarrow 5$ ,  $8 \rightarrow 6$ , (w kierunku angielsko-polskim)  $1 \rightarrow \text{NULL}$ ,  $2 \rightarrow 3$ ,  $3 \rightarrow 4$ ,  $4 \rightarrow 5$ ,  $5 \rightarrow 7$ ,  $6 \rightarrow 8$ . Przecięcie to:
- 2-3, 3-4, 4-5, 5-7, 6-8
  - 2-3, 3-4, 5-7, 6-8
  - 2-2, 2-3, 3-4, 4-5, 5-7, 6-8
  - 5-7, 6-8
17. Symetryzacja dopasowań. Dla zdań „Mieszkańcy całej ziemi mieli jedną mowę” i „And the whole earth was of one language” mamy funkcję dopasowania: (w kierunku polsko-angielskim)  $1 \rightarrow \text{NULL}$ ,  $2 \rightarrow 2$ ,  $3 \rightarrow 2$ ,  $4 \rightarrow 3$ ,  $5 \rightarrow \text{NULL}$ ,  $6 \rightarrow \text{NULL}$ ,  $7 \rightarrow 5$ ,  $8 \rightarrow 6$ , (w kierunku angielsko-polskim)  $1 \rightarrow \text{NULL}$ ,  $2 \rightarrow 3$ ,  $3 \rightarrow 4$ ,  $4 \rightarrow 5$ ,  $5 \rightarrow 7$ ,  $6 \rightarrow 8$ . Suma to:
- 2-2, 2-3, 3-4, 4-5, 5-7, 6-8

- 5-7, 6-8
- 2-3, 3-4, 5-7, 6-8
- 2-3, 3-4, 4-5, 5-7, 6-8

18. Symetryzacja dopasowań. Dla zdań „Mieszkańcy całej ziemi mieli jedną mowę” i „And the whole earth was of one language” mamy funkcję dopasowania: (w kierunku polsko-angielskim)  $1 \rightarrow \text{NULL}$ ,  $2 \rightarrow 2$ ,  $3 \rightarrow 2$ ,  $4 \rightarrow 3$ ,  $5 \rightarrow \text{NULL}$ ,  $6 \rightarrow \text{NULL}$ ,  $7 \rightarrow 5$ ,  $8 \rightarrow 6$ , (w kierunku angielsko-polskim)  $1 \rightarrow \text{NULL}$ ,  $2 \rightarrow 3$ ,  $3 \rightarrow 4$ ,  $4 \rightarrow 5$ ,  $5 \rightarrow 7$ ,  $6 \rightarrow 8$ . Efekt GROW-DIAG-FINAL to:

- 2-3, 3-4, 5-7, 6-8
- 2-2, 2-3, 3-4, 4-5, 5-7, 6-8
- 2-3, 3-4, 4-5, 5-7, 6-8
- 2-2, 2-3, 3-4, 5-7, 6-8

19. Rozpatrujemy parę zdań „Ala ma kota”=„Alice has a cat” i dopasowanie 1-1, 2-2, 3-4. Która z podanych fraz zostanie pozyskana:

- „a”=„ma”
- „a”=„NULL”
- „ma kota”=„has a cat”
- „ma kota”=„has a”

20. Rozpatrujemy parę zdań „jest to książka”=„this is a book” i dopasowanie  $1 \rightarrow 2$ ,  $2 \rightarrow 1$ ,  $1 \rightarrow 3$ ,  $3 \rightarrow 4$ . Która z podanych fraz NIE zostanie pozyskana:

- „jest to książka”=„this is a book”
- „jest”=„is a”
- „jest to”=„is a”
- „książka”=„book”

21. Rozpatrujemy parę zdań „Ala ma kota”=„Alice has a cat” i dopasowanie 1-1, 2-2, 3-4. Ile fraz zostanie w ogóle zostanie pozyskanych?

- 9
- 7
- 8
- 6

22. Niech  $t1 =$  „od czasu do czasu”,  $t2 =$  „lufcik sfera sfera purpurowymi”,  $t3 =$  „znaleziono czerwony wózek dziecięcy”, należałoby oczekiwać, że dla dobrego modelu języka LM (dla języka polskiego):

- $p_{LM}(t1) < p_{LM}(t3) < p_{LM}(t2)$
- $p_{LM}(t1) > p_{LM}(t2) \approx p_{LM}(t3)$
- $p_{LM}(t1) > p_{LM}(t3) > p_{LM}(t2)$
- $p_{LM}(t1) > p_{LM}(t2) > p_{LM}(t3)$

23. Przyjmijmy prostą estymację prawdopodobieństw unigramów wykonanych na jakimś typowym polskim korpusie. Jaka estymacja jest najbardziej prawdopodobna dla najczęstszego polskiego słowa („w”):

- $\approx 0.01$
- $\approx 1.1$
- $\approx 0.5$
- $\approx 0.95$

24. Załóżmy, że wygenerowano model języka na podstawie zdania „róża jest różą jest różą” (bez wygładzania). Ile wynosi  $P(\text{jest}|\text{róża})$ ? Uwzględnić znaczniki początku i końca.

- 1
- $1/2$
- $1/3$
- $2/3$

25. Przyjmijmy prostą estymację prawdopodobieństw unigramów wykonanych na jakimś typowym polskim korpusie. Jaka wartość logprob (przy podstawie 10) jest najbardziej prawdopodobna dla wyrazu takiego jak „purpurowy”

- $\approx 0.0005$
- $\approx -0.5$
- $\approx 1.0005$
- $\approx -5$

26. Przyjmijmy dobry model języka polskiego. Jaka wartość cross-entropii dla sensownego korpusu ewaluacyjnego jest możliwa?

- $\approx 1$
- $\approx 0$
- $\approx 2$
- $\approx 8$

27. Jeśli cross-entropii wynosi 8, to perplexity wynosi:

- 256
- 3
- 0.02
- 0.125

28. Kiedy korpus ewaluacyjny rośnie (zaś model języka pozostaje bez zmian):

- perplexity wzrasta liniowo
- perplexity pozostaje mniej więcej takie same
- perplexity spada liniowo
- perplexity wzrasta wykładniczo

29. Jaką średnią liczbę wystąpień w odłożonym korpusie spodziewałbyś się dla wyrazów, które wystąpiły jeden raz (hapax legomena) w podstawowym korpusie?

- $\approx 2$
- $\approx 0.01$

- $\approx 0.5$
- $\approx 1.1$

30. Kiedy (sensowny) model jest tworzony na coraz większym korpusie (zaś korpus ewaluacyjny pozostaje bez zmian):

- perplexity pozostaje mniej więcej takie same
- perplexity zachowuje się w sposób nieprzewidywalny
- perplexity spada
- perplexity wzrasta

31. Kiedy model jest tworzony na tym samym korpusie, ale ma coraz większy rząd (coraz dłuższe n-gramy są brane pod uwagę), zaś korpus ewaluacyjny pozostaje bez zmian:

- perplexity wzrasta
- perplexity zachowuje się w sposób nieprzewidywalny
- perplexity pozostaje mniej więcej takie same
- perplexity spada

32. Związek między jakością modelu a perplexity:

- Im lepszy model, tym niższe perplexity na odłożonym korpusie.
- Im lepszy model, tym wyższe perplexity na odłożonym korpusie.
- Im lepszy model, tym bardziej perplexity przybliży się do 100 procent.
- Nie można stosować perplexity do oceny jakości modelu.

33. Model log-liniowy to liniowa kombinacja logarytmów

- dwóch feature functions (które niekoniecznie reprezentują prawdopodobieństwo)
- dwóch feature functions (które muszą reprezentować prawdopodobieństwo)
- dowolnej liczby feature functions (które niekoniecznie reprezentują prawdopodobieństwo)
- dowolnej liczby feature functions (które muszą reprezentować prawdopodobieństwo)

34. Jakie dopuszczalne wartości przyjmuje logarytm prawdopodobieństwa (logprob)?

- Od 0 do plus nieskończoności.
- Od 1 do plus nieskończoności.
- Od minus nieskończoności do 0.
- Od 0 do 1.

## KLUCZ

1) 4, 2) 4, 3) 4, 4) 3, 5) 3, 6) 3, 7) 1, 8) 2, 9) 4, 10) 1, 11) 3, 12) 4, 13) 3, 14) 1, 15) 4, 16) 2, 17) 1, 18) 2, 19) 3, 20) 3, 21) 1, 22) 3, 23) 1, 24) 2, 25) 4, 26) 4, 27) 1, 28) 2, 29) 3, 30) 3, 31) 4, 32) 1, 33) 3, 34) 3, 35) ,